(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 253 516 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
30.10.2002 Bulletin 2002/44

(51) Int Cl.7: G06F 9/50

(21) Application number: 02009207.8

(22) Date of filing: 24.04.2002

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 25.04.2001 US 843426

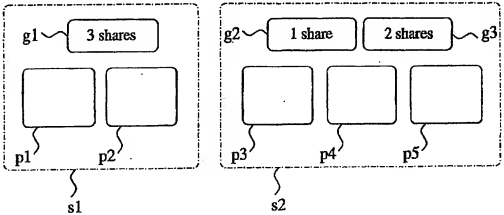
(71) Applicant: Sun Microsystems, Inc.
Palo Alto, California 94303 (US)

(72) Inventors:
- Dorofeev, Andrei V.
Sunnyvale, CA 94086 (US)
- Tucker, Andrew G.
Menlo Park, CA 94025 (US)

(74) Representative: HOFFMANN - EITLE
Patent- und Rechtsanwälte
Arabellastrasse 4
81925 München (DE)

(54) Apparatus and method for scheduling processes on a fair share basis

(57) Described is a scheduling system that provides allocation of system resources of one or more processor sets among groups of processes. Each of the process groups is assigned a fixed number of shares, which is the number that is used to allocate system resources among processes of various process groups within a given processor set. The described fair share scheduler considers each processor set to be a separate virtual computer. Different process sets do not share processes, a particular process must execute on a single processor set. In another embodiment of the invention, each process group could be given a separate number of shares for each processor set. Percentage of the resources of the specific processor set allocated to processes of a process group is calculated as a ratio of the shares of the process group on the processor set to the total number of shares of active process groups operating in that set. The process group is considered active on a processor set, if that processor set executes at least one process of that process group.



P_{1-5} = 20% (1/5th) of the system resources.

Fig. 1

EP 1 253 516 A2

Printed by Jouve, 75001 PARIS (FR)

[0015] FIG. 1 illustrates exemplary allocation of system resources according to the inventive concept.

DETAILED DESCRIPTION

[0016] To overcome the limitations described above, and to overcome other limitations that will become apparent upon reading and understanding the present specification, apparatus, methods and articles of manufacture are disclosed that allocate a percentage of system resources among process groups in a computer system having one or more processor sets.

[0017] According to the inventive method, processes are combined into process groups based on a pre-defined criteria. Each of the process groups is assigned a number of shares representing relative importance of the group within its processor set. The inventive system allocates the system resources of the processor sets to process groups according to the number of shares assigned to a particular process group and the total number of shares of all active process groups in the processor set. A process group is considered active on a processor set if there is at least one process of this process group executing on that processor set.

[0018] Fair share scheduling is a way to assign a particular process a fixed share of CPU resources. The term share may be used to describe the relative importance of one workload versus another.

[0019] According to one of the aspects of the present invention, various processes in the system are combined into one or more process groups. These process groups users are assigned a number of shares which represent relative importance thereof. This is a way to guarantee application performance by explicitly allocating shares (or percentage) of system CPU resources among competing workloads. Note that total number of shares assigned to all process groups need not be 100. Furthermore, to obtain the percentage of the system CPU resources available to a process group at each given moment of time, a total number of shares allocated to that process group must be divided by the total number of shares possessed by all currently active process groups. A process group is considered active when it has at least one running or runnable process. Indeed, to ensure the complete, or 100% utilization of the system, only process groups which have executing processes at a particular time should be given share of the CPU usage. Note that such active process groups are searched across the entire system. At any given time, the percentage of the CPU allocated to a particular process group depends on the number of shares owned by all other active process groups in the system, or the process groups that have at least one executing process in the system. Therefore, in a system where processes are combined into process groups based on user id, any new logged user with a given number of shares can decrease the CPU percentage of all other actively running users.

[0020] Modern operating systems, such as a Solaris Operating System distributed by Sun Microsystems, Inc. of Palo Alto, California have a concept of processor sets. Processor set concept applies to multiple processor computers and allows the binding of one or more processors into groups of processors. Processors assigned to processor sets will run only processes that have been bound to that processor set. In other words, the aforementioned processor set is essentially a virtual single- or multi-processor computer system within a physical computer, which has its own set of running processes. The concept of processor set is especially helpful, for example, when certain important process need to be provided with a separate one or more processors. For example, in a computer system providing services to http clients, a separate processor set can be allocated to running a web server, while all other processes can be executed on a second, separate processor set. In this case, the amount of CPU resources allocated to the web server will not depend on the other processes executing in the system.

[0021] However, when the aforementioned processor sets are used in conjunction with the conventional fair share scheduler, the performance of processes running on one processor set may be impacted by the work performed by processes running on another processor set, which is an undesirable effect.

[0022] The reason why the existing fair share scheduler does not work satisfactorily with processor sets is because that total number of shares for all active process groups is calculates across the entire system, when in fact it should be calculated only within boundaries of the current processor set. If the total number of shares is kept separate for each processor set, then the CPU allocation for a given process group will only depend on other process groups who have their active processes on the same processor set. The work done on other processor sets will be unaffected. This is more intuitive behavior of such configurations than what it has been in the past.

[0023] The inventive fair share scheduler will now be described in detail. According to the inventive concept, various processes in a computer system are combined into process groups. Each of these process groups is assigned a fixed number of shares, which is the number that represents relative importance of process groups. The number of shares of a process group is used to allocate system resources among processes of that process group executing within a predetermined processor set, in the manner described in detail below. Specifically, the inventive fair share scheduler considers each processor set to be a separate virtual computer. Different processor sets do not share processes, in other words, a process must execute on a single processor set.

[0024] In one embodiment of the invention, each process group is given the same number of shares for all processor set. It should be noted that even if process group has zero shares, processes of this process group

improved performance characteristics.

[0037] According to another embodiment, a computer system may be provided comprising at least a central processing unit and a memory, said memory storing a program for allocating a percentage of system resources among process groups in a computer system, said computer system comprising at least one central processing unit, said at least one central processing unit combined into at least one processor set, said program comprising instructions for assigning each of said process groups a number of shares for each or said at least one processor set allocating said system resources of each of said at least one processor set to each of said process groups according to the number of shares assigned to said each of said process groups.

[0038] Further, it is noted that a computer-readable medium may be provided having a program embodied thereon, where the program is to make a computer or a system of data processing devices to execute functions or operations of the features and elements of the above described examples. A computer-readable medium can be a magnetic or optical or other tangible medium on which a program is recorded, but can also be a signal, e.g. analog or digital, electronic, magnetic or optical, in which the program is embodied for transmission. Further, a computer program product may be provided comprising the computer-readable medium.

Claims

1. A method for allocating a percentage of system resources among process groups in a computer system, said computer system comprising at least one central processing unit, said at least one central processing unit combined into at least one processor set, said method comprising:

   a. assigning each of said process groups a number of shares for each or said at least one processor set;

   b. allocating said system resources of each of said at least one processor set to each of said process groups according to the number of shares assigned to said each of said process groups.

2. The method of claim 1, wherein said system resources of each of said at least one processor set are allocated based on a number of shares of all active groups within said each of said at least one processor set.

3. The method of at least one of the claims 1 and 2, wherein said percentage of said system resources is calculated based on a ratio of the number of shares assigned to said each of said process groups to the a number of shares of all active groups within said each of said at least one processor set.

4. The method of at least one of the claims 1 to 3, wherein each of said process groups includes only one process.

5. A computer readable medium embodying a program for allocating a percentage of system resources among process groups in a computer system, said computer system comprising at least one central processing unit, said at least one central processing unit combined into at least one processor set, said program comprising:

   a. assigning each of said process groups a number of shares for each or said at least one processor set;

   b. allocating said system resources of each of said at least one processor set to each of said process groups according to the number of shares assigned to said each of said process groups.

6. The computer readable medium of claim 5, wherein said system resources of each of said at least one processor set are allocated based on a number of shares of all active groups within said each of said at least one processor set.

7. The computer readable medium of at least one of the claims 5 and 6, wherein said percentage of said system resources is calculated based on a ratio of the number of shares assigned to said each of said process groups to the a number of shares of all active groups within said each of said at least one processor set.

8. The computer readable medium of at least one of the claims 5 to 8, wherein each of said process groups includes only one process.

9. A program having instructions adapted to make a computer carry out the method of at least one of the claims 1 - 4.

10. A scheduler for allocating a percentage of system resources among process groups in a computer system having at least one central processing unit, said at least one central processing unit combined into at least one processor set, the scheduler comprising:

    means for assigning each of said process groups a number of shares for each or said at least one processor set; and

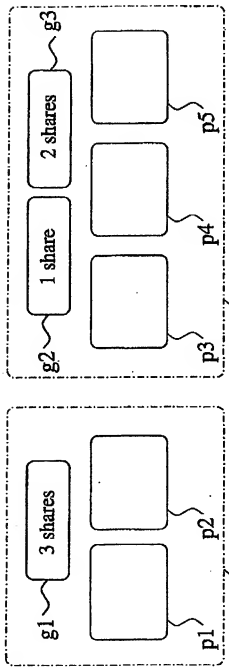

g3
2 shares
g2
1 share
p5
p4
p3
s2
g1
3 shares
p2
p1
s1
P1–5 = 20% (1/5th) of the system resources.

Fig. 1